



## Cost damping and functional form in transport models

Rich, Jeppe; Mabit, Stefan Lindhard

*Published in:*  
Transportation

*Link to article, DOI:*  
[10.1007/s11116-015-9628-8](https://doi.org/10.1007/s11116-015-9628-8)

*Publication date:*  
2016

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Rich, J., & Mabit, S. L. (2016). Cost damping and functional form in transport models. *Transportation*, 43(5), 889–912. <https://doi.org/10.1007/s11116-015-9628-8>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Cost damping and functional form in transport models

Jeppe Rich<sup>✉</sup>, Stefan L. Mabit

*Department of Transport, Technical University of Denmark*

*Bygningstorvet 16B, 2800 Kgs. Lyngby*

*Email: [jr@transport.dtu.dk](mailto:jr@transport.dtu.dk)*

*Phone: +45 45251536*

*Fax: +45 45936533*

## Abstract

Transport models allowing for cost damping are characterised by marginally decreasing cost sensitivities in demand. As a result, cost damping is a model extension of the simple linear-in-cost model requiring an appropriate non-linear link function between utility and cost. The link function may take different forms and be represented as a non-linear-in-parameter form such as the well-known Box-Cox function. However, it could also be specified as non-linear-in-cost but linear-in-parameter forms, which are easier to estimate and improve model fit without increasing the number of parameters. The specific contributions of the paper are as follows. Firstly, we discuss the phenomenon of cost damping in details and specifically why it occurs. Secondly, we provide a test of damping and an easy assessment of the (linear) damping rate for any variable by estimating two auxiliary linear models. This turns out to be an important guidance as the damping rate largely dictates which link functions are appropriate for the data. Thirdly, inspired by the Box-Cox function, we propose alternative linear-in-parameter link functions, some of which are based on interpolation of approximate Box-Cox end points, and others which are inspired by Taylor Expansions. The different functions are tested in simulation experiments and subsequently in a large-scale demand model based on more than 22,000 revealed preference observations. It is concluded that the use of properly specified linear-in-parameter functions gives good data fit and sometimes even outperforms the Box-Cox functions without increasing the number of parameters.

*Keywords: cost damping, functional form, Box-Cox, travel demand, discrete choice analysis.*

# 1 Introduction

An important element of estimating travel demand models is to find the functional form that explains the underlying data generating process in the best possible way. In fact, the increase in model fit from using a proper functional form usually outweighs most other model-related choices such as accounting for heterogeneity and correcting for correlation in the error terms. However, the choice of functional form is not only important with respect to model fit but also because it affects all properties of the model from direct elasticities and cross elasticities to the derived value of time (VoT). If the model is concerned with long-distance trips the model needs to explain observations across a wide range of travel distances, which can be challenging. Therefore the importance of functional form tends to increase with the distance domain of the model.

A search in the literature reveals three approaches to model functional form of a single variable: piecewise linear approximation (splines), power series expansions, and Box-Cox transformations. The first approach is discussed in Ben-Akiva and Lerman (1985) in a logit modelling framework and more recently in Pinjari and Bhat (2006) in a mixed logit framework. Power series expansions are discussed in Ben-Akiva and Lerman (1985) and have been applied in all types of transport modelling contexts as a simple test of the base model that is linear in variables. The work on Box-Cox transformations in the travel demand literature goes back to the early work by Gaudry and Wills (1978) and Hensher and Johnson (1981). The focus of the present paper is dedicated to the latter. Gaudry (2010) summarises the early findings and provides many arguments why the Box-Cox model is useful for modelling of travel demand. Applications are found in Gaudry et al. (1989), Ben-Akiva et al. (1987), Mandel et al. (1994) and Lapparent and de Palma (2002).

Hyman (2007) argues that demand models, in order to be realistic, should comply with at least three properties: 1) decreasing sensitivity to generalised costs for longer than average trips, 2) increasing VoT for longer trips, and 3) monetary costs should enter in the generalised form in a logarithmic form. While the first two properties are intuitive and confirmed elsewhere (Jara-Diaz, 2007) it is important to stress that cost damping (of which property 3) is a specific case) is primarily a phenomenon that plays out on longer distances. For many urban models where the trip distances are short or moderate, the damping is usually low and may even be negative. Daly (2010), in his work on cost damping in travel demand models and in line with the first two properties discussed by Hyman (2007), argues that a desirable model should:

- i) be RUM consistent in the sense that the model should be consistent with random utility maximisation (RUM) such that the gradient of the utility function with respect to cost is equal or less than zero for the entire domain of the model.
- ii) allow for cost damping (decreasing sensitivity to cost and/or time), and
- iii) allow for increasing VoT with distance.

The first property is the softest as it establishes a general type of RUM consistency. The second property should not be considered a restriction but rather a relaxation of the linear-in-attribute model where the econometric specification of the model is extended to allow for a more flexible curvature. In doing so we allow for a parameterisation of the “de-acceleration” of slope when the distance increases. Although there are no strict theoretical arguments for cost damping there is strong empirical evidence in favour of models

applying cost-damping as discussed in Daly (2010). In Section 2 we will elaborate further on this and argue that cost damping arises from several confounding processes and should be seen, primarily, as a sound econometric extension of the linear model. The third property, we believe, should be generalised slightly as to say,

- iii) allow for different degrees of damping for different attributes (cost and time) with distance.

Empirically, there is a tendency that cost is more damped than time, which is why Daly (2010) refers to increasing VoT as a function of distance. However, this observation is mainly true for daily trips where time constraints play a role. For longer trips and in particular for vacation trips, this is not necessarily the case as the trip in itself may be considered part of the journey. To illustrate the link between the VoT and cost damping consider a simple utility function where cost  $c$  is modelled as  $f(c, \beta_c)$  and time  $t$  as  $g(t, \beta_t)$ . VoT is given by  $VoT = \frac{\partial g(t, \beta_t)}{\partial t} / \frac{\partial f(c, \beta_c)}{\partial c}$ . If the damping in  $c$  is higher than in  $t$  we will have that  $\frac{\partial g(t)}{\partial t}$  dominates  $\frac{\partial f(c)}{\partial c}$  for longer distances. Hence, as distance increases,  $VoT$  will increase as well. In a linear version of  $g(t, \beta_t)$  and  $f(c, \beta_c)$ ,  $VoT$  is simply the ratio between  $\beta_t$  and  $\beta_c$ . So there is no damping. In a version with  $g(t, \beta_t) = \beta_t \ln(t)$  and  $f(c, \beta_c) = \beta_c \ln(c)$  we have  $VoT = \beta_t c / \beta_c t$ . This is a version in which both time and cost are “maximally damped” as will be considered in more detail in the subsequent sections. In the current paper, we will explore damping in a generalised cost context as we maintain a fixed VoT formulation. It does not affect the generality of the results as our main objective is to propose a set of functions that will apply in any context including a context, where we have separate damping functions for cost and time. However, it means that a more flexible specification as represented by iii) is worth while considering if supported by data and that possible challenges that arise from the estimation of such forms are while worth considering for the future as also pointed out in Section 5.

Generally, the three properties can be used to remove specific functional forms as potential candidates in a demand model. A candidate that fulfils all the properties is the Box-Cox transformation. However, due to its non-linearity it has not been widely used in (practical) large-scale modelling. Moreover, although being flexible and generally quite successful in many applications (e.g., Gaudry, 2010) it should be stressed that the Box-Cox transformation in itself represents an approximation to an underlying unknown data generating process. Hence, there is no guarantee (or theoretical evidence) supporting the Box-Cox model for other functions. For the estimation of Constant Elasticity of Substitution (CES) functions, however, the Box-Cox model will have the advantage of being monotonic as well as having continuity for lambda equal to zero (Gaudry, 2010). In the context of discrete choice this is less important as the models are typically restricted to Box-Cox parameters in the unit interval. In that case all of the models presented in this paper will preserve the order of the data.

In this study we investigate how the choice of functional form affects travel demand models and we give numerous examples of useful forms other than the Box-Cox model. To focus the analysis in the paper we limit the analysis and the simulation experiments to assess the functional form of generalised travel time variables and not separate time and cost terms as well as the corresponding VoT. However, as damping and non-linearity in general are found not only in specifications with generalised time and cost, the findings in the paper are also relevant to cases where VoT is estimated. Hence, the definitions and the estimation of the damping rate considered in Section 2 as well as the different functional forms considered in Sections 3 and 4 are also relevant for cases where VoT is estimated. Even so, as many practical models world-wide

have applied generalised cost or time specifications, the investigation of functional form in a generalised time context is also of practical relevance. For a more elaborate discussion of the concept of generalised time and cost we refer to Wardman (2008).

The remainder of the paper is organised as follows. First, we discuss cost damping in general and functional form in Section 2 where we also advice on how to assess the amount of damping in a data set. In Section 3, we suggest various approximations to the Box-Cox model and investigate by means of simulated data how these can be used to model data both when the true model is a Box-Cox model and when it is not. Section 4 illustrates how the various functional forms can be used in a large-scale demand model and the effects of functional form on model fit, elasticities, and recuperation of the distance distribution in the data. The final Section 5 concludes the paper.

## **2 Functional form and cost damping**

Cost damping has been observed repeatedly in the empirical literature. This is particular true for long-distance models, see e.g. Rohr et al. (2010), Börjesson (2010) and Rich and Mabit (2012). As mentioned in Section 1 we will consider damping in a broader context as a property that relates to, e.g. monetary cost, travel time attributes, and generalised cost. For the attributes being considered, it is common to find cost damping although the rate may vary depending on which attribute is considered. Experiments carried out in connection with the Danish National Transport Model (NTM) show that cost damping is a phenomenon that is also very much present in travel time attributes. A relevant question is therefore to ask why cost damping occurs. Although there may be other causes, we believe that cost damping is a phenomenon that arises from at least the following processes.

Firstly, it is possibly related to human preferences as represented by the utility function not being entirely linear with respect to increasing costs. One perception is that all trips have some sort of inherent start-up costs. This means that the perceived cost is higher for shorter trips and less for longer trips which in turn may show up as damping.

Secondly, cost damping may occur as an effect of unobserved attributes, e.g. car occupancy rate. Occupancy rate is in general increasing with distance. If a model is too simplistic and does not capture occupancy rate effects properly, it will not capture that costs are relatively deflated as distance increases whereas the time sensitivity remains constant. In models where car costs are attached to the driver, e.g. no cost sharing, this would most likely result in cost damping. This is because the higher occupancy rate for longer distances tends to deflate the individual cost all other things being equal.

A third possible reason for damping may be heterogeneity in the error term. As distance increases the error term accumulates more noise and the variance goes up. When the variance increases, the scale goes down and will consequently cause model sensitivity to decrease. This is discussed in more detail in Daly and Carrasco (2009).

A fourth possible reason may be selection bias. The higher travel cost, the more valuable is the time at the destination.

Whatever the reason, not correcting for cost damping may lead to bias in the model estimation as well as in the sensitivity of the model. In our view, the simplest and possibly most efficient strategy<sup>1</sup> is to consider cost damping as a functional form issue and this will therefore be the departure point of the paper.

## 2.1 Definition of damping

As stated in Section 1 damping is defined as “decreasing relative sensitivity to cost or time”. In what follows we will skip the distinguished reference to cost or time and simply refer to “cost” as a common term. There are natural limits to how much and how little damping a model should have. These limits can be assessed by considering the sensitivity to increasing costs. Consider the logit model for two simple utility functions  $U_i = ax_i + \epsilon_i$  and  $V_i = b \ln x_i + \epsilon_i$  where  $\epsilon_i$  are error terms that are identically and independently Gumbel distributed. If  $P_i$  represents the logit probability function, the direct elasticity of these two models is given in (1) below.

$$(1) \quad E_{U,i} = ax_i(1 - P_i) \quad \text{and} \quad E_{V,i} = b(1 - P_i)$$

In the linear case, there is proportionality between the elasticity and the attribute, and the sensitivity is equal to  $ax_i$  as the cost goes to infinity. This will be referred to as “minimum cost damping” as the relative effect of cost is not reduced as the underlying attribute increases. Having a model with  $ax_i^\delta$  with  $\delta > 1$  would in principle be possible, however, it would lead to an aggressive scaling when  $x_i$  is large. Usually  $\delta > 1$  does not apply to long-distance travel although it may be the case for shorter distances. On the other hand, the logarithmic case may be referred to as “maximum cost damping” as the elasticity approaches  $b$  as the distance goes to infinity and hence is invariant to the cost level. If we go beyond the maximum damping case it usually implies a violation of the RUM assumption (the first property in Section 1). The above discussion can be operationalised through Definition 1 below.

---

**Definition 1:** Let  $x$  represent travel costs and  $f(x) = a \ln x + bx$  a logarithmic and linear transformation of  $x$  which is introduced in the travel demand model. We will say that  $x$  is

- damped if  $a < 0, b < 0$  and both are statistically significant.
  - maximally damped if  $a < 0$  and statistically significant, and  $b \approx 0$ , i.e. not significantly different from 0.
  - minimally damped if  $b < 0$  and statistically significant, and  $a \approx 0$ , i.e. not significantly different from 0.
- 

<sup>1</sup> Correcting for the functional form will repair not only the non-linearity in the preference structure, but also (at least partly) the effect that heterogeneity may have on model sensitivity. On the other hand, accounting for heterogeneity will not correct for non-linearity in the preference structure.

It is important to note that this definition is not equivalent to saying that every damped process should follow the  $a \ln x + bx$  form. Rather it is a technical definition stating that if, in a given data generating process, the log-term  $a \ln x$  and the linear term  $bx$  become significant with a negative parameter, then the data generating process will be partly damped. A visualisation of Definition 1 is shown below in Figure 1.

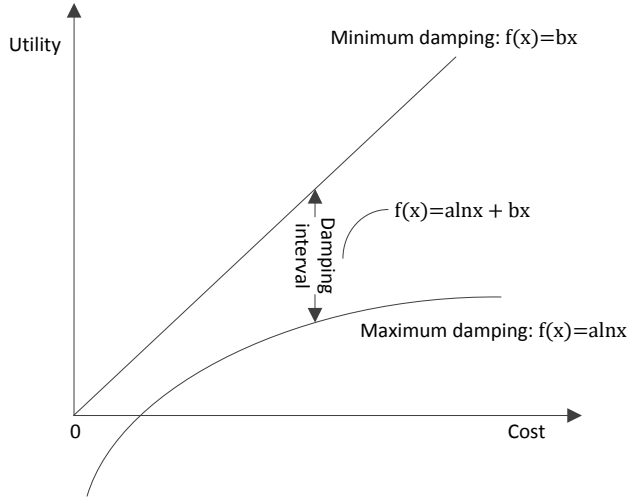


Figure 1: Illustration of minimum and maximum damping.

There could be other possible definitions of damping, which could be anchored in the Box-Cox model, however Definition 1 is intended as an operational and easy-to-assess damping measure that does not require non-linear models to be estimated.

## 2.2 Assessing the amount of damping

A very interesting question is whether we can define a measure for the amount of damping in a given data generating process. This is considered in Definition 2 below.

---

**Definition 2:** In a dampened data generating process  $x$  (refer to Definition 1) we will refer to  $\mu \in [0,1]$  as the “linear damping rate” of  $x$ . Consider the two auxiliary models represented by:

$$(2) \quad \begin{array}{ll} M1 & f(x) = z\theta + a_2x + b \ln x \\ M2 & f(x) = z\theta + a_1x \end{array}$$

The “linear damping rate”  $\mu$  of  $x$  is defined by

$$(3) \quad \hat{\mu} = 1 - \frac{\hat{a}_2}{\hat{a}_1}$$

If  $\mu = 1$ ,  $x$  is said to be maximally damped and if  $\mu = 0$ ,  $x$  is said to be minimally damped.

---

The linear damping rate is estimated under the  $H_0$  hypothesis with minimal damping and under the assumption that the functional form is linear. In the case of minimal damping the two models should be identical and  $\hat{a}_1 = \hat{a}_2$ . If there is maximum damping  $\hat{a}_2 \rightarrow 0$  and the logarithmic term in  $M1$  will overtake the specification. As mentioned, the specification is only correct under  $H_0$  or in an infinitesimal region of  $H_0$ . So the power of the test (of damping) and thereby the precision of the estimate of  $\mu$  will gradually decline as the amount of damping increases.

In the estimation of  $a_1$  in model  $M2$  we recommend replacing  $\theta$  with the parameter vector  $\hat{\theta}$  from Model  $M1$ . In doing so we assure that the  $\theta$ -term is identical across models  $M1$  and  $M2$ . So first we estimate a fully parameterised model ( $M1$ ) and by re-using everything except for the linear parameter for  $x$ , we avoid that correlation between this term and other terms affects the linear parameter estimate in the second model.

It should be noted that the above diagnosis (Definition 2) applies only to data where the damping rate is between 0 and 100% and where the lower damping bound can be represented by a linear term. If this is not the case, the power of the test will decline but may still provide valuable information. However, if the damping approaches the lower bound (the linear case) the damping may be more aggressive than the linear term and should be modelled as a power function with powers above 1. This corresponds to a Box-Cox parameter above unity and can be the case for urban models where the travel distances are short or moderate.

At this stage it is natural to introduce the Box-Cox in which the amount of damping is represented by the  $\lambda$  parameter.

$$(4) \quad BC(x; \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \in ]0,1] \\ \ln(x) & \lambda = 0 \end{cases}$$

The linear and logarithmic cases are nested within the BC model and emerge as  $\lambda \rightarrow 1$  and  $\lambda \rightarrow 0$ , respectively. These limits then correspond to the “minimum” and “maximum” damping definitions in Definition 1 as discussed above. It should be noted that in general, the Box-Cox is not necessary restricted to  $\lambda \in ]0,1]$ , so we consider a restricted version of the Box-Cox model in this paper.

Both  $\hat{\mu}$  and  $1 - \lambda$  represent a measure of cost damping and they both operate on the unit interval. In Table 1 below we have compared the two measures of damping in a Monte Carlo simulation experiment where a true Box-Cox data generating process has been simulated for different values of  $1 - \lambda$ .

True $1 - \lambda$	$\hat{\mu}$	Abs difference	% deviation
0.98	0.97	0.01	28.5%
0.9	0.87	0.03	26.1%
0.8	0.75	0.05	23.1%
0.7	0.64	0.06	20.1%
0.6	0.53	0.07	16.9%
0.5	0.43	0.07	14.2%
0.4	0.33	0.07	11.2%
0.3	0.24	0.06	8.4%



0.2	0.16	0.04	5.5%
0.1	0.08	0.02	2.7%
0.02	0.01	0.01	0.5%

Table 1: Comparison of  $\hat{\mu}$  and  $\lambda$  in a true simulated Box-Cox data generating process. The deviation is calculated as  $(1 - \hat{\mu} - \lambda)/\lambda$ .

The first thing to note is that  $1 - \lambda$  and  $\hat{\mu}$  measure different types of damping and may not be identical although they will be similar under  $H_0$ . Also, as the BC function approaches the log when  $\lambda \rightarrow 0$  we will have  $\hat{\mu} = 1 - \lambda$  in the limit. As the test is formulated under  $H_0$ , the power of the test will decline as we move away from  $H_0$  and so will the relative precision of  $\hat{\mu}$  as it appears.

What happens here is that the Box-Cox model measures damping in terms of the Box-Cox functions whereas  $\hat{\mu}$  measures damping in a linear way. This is illustrated below in Figure 2.

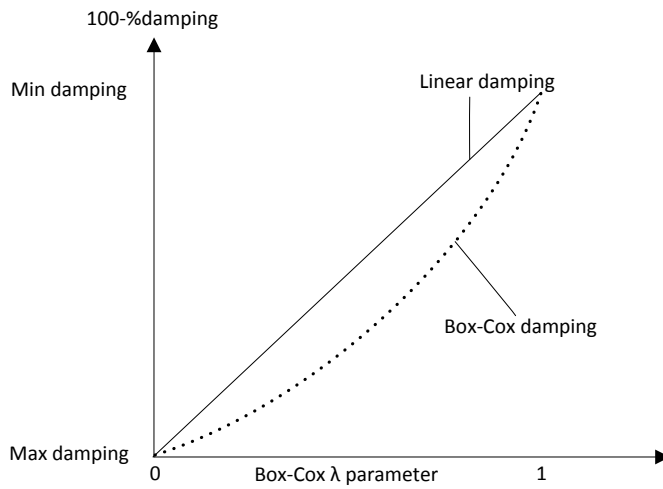


Figure 2: Damping in the Box-Cox model and in the linear model.

An exact relationship between  $\mu$  and  $\lambda$  depends on the underlying unobserved data generating process and cannot be explicitly stated. However, it is clear that the Box-Cox function is strictly convex in  $\lambda$  for  $x > 1$  and that  $\mu < 1 - \lambda$ .

### 3 Approximating the Box-Cox

The Box-Cox model provides a flexible family of functional forms with all the desirable properties discussed in the introduction. Hence, it is a very natural place to start if we are looking for candidate functions allowing for cost damping. Estimating Box-Cox functions, however, can be challenging for large scale modelling due to its non-linear-in-parameter form and several packages such as ALOGIT and STATA do not currently facilitate Box-Cox functions in the estimation of the discrete choice models. SAS/QLIM and Biogeme do support Box-Cox functions, however, for large samples (>20.000 observations) and for large choice sets (>5000 alternatives) the introduction of non-linear forms is still computationally cumbersome. In any case, for testing and screening purposes, it is useful to consider approximations. As a result, we will look for appropriate Box-Cox approximations expressed as linear-in-cost functions.

In the following we will consider linear-in-parameter approximations of the Box-Cox model based on the findings in Section 2. A natural approximation is to define the approximating function as a mixture of the linear and logarithmic curve. That is;

$$(5) \quad L(x; a, b) = ax + b \ln x$$

This form is suggested by Hyman (2007) and discussed in Daly (2010). The form can be seen as a special case of two Box-Cox functions. Generally, if in the underlying Box-Cox functions the regression terms have the same sign, (5) will lead to an acceptable approximation of the Box-Cox process as the underlying form is monotone and has not a minimum or maximum (Gaudry et al. 2000). If that is not the case the first requirement in Section 1 will be violated. In that sense we may also see equation (5) as an approximation which imposes structure.

The function has been used in several large-scale models. In Sweden, it was used in Beser and Algers (2002) and more recently in a report by WSP (WSP, 2011) on model development for high speed rail modeling. In the UK, it was used in Rohr et al. (2010). A problem with (5) is that for certain data generating processes, correlation between the linear and logarithmic parts may cause one of the parameters to be positive. Hence, the approximation violates the RUM condition discussed in Section 1. An alternative one-parameter formulation can be attributed to Daly (2012), which is often referred to as the Gamma form.

$$(6) \quad LL_B(x; \gamma) = \gamma x + (1 - \gamma) \ln x - \gamma$$

The model in (6) uses only one parameter and is by construction specified so that

- at  $x = 1$  both  $BC(x, \gamma)$  and  $LL_B(x; \gamma)$  are zero and with unity slope.
- at  $x = 1$  the second derivative for both specifications is  $(\gamma - 1)$ .
- The functions are always equal if  $\gamma = 0$  and  $\gamma = 1$ .

Provided  $x$  is close to 1 the approximation in (6) yields a good approximation to the Box-Cox model. Although this can to some extent be accomplished by dividing by the mean of  $x$ , the fit may not be particularly good in the tail of the distribution if the domain of  $x$  is wide.

Clearly, the objective of these models has not been to mimic the Box-Cox model but to model a latent data generating process which obviously may be quite different from Box-Cox data. Irrespectively whether the model in (5) has been considered as an approximation of the Box-Cox or not, the answer to a good Box-Cox approximation “lies in its genes” in the sense that interpolation is a good strategy.

More specifically, the logarithmic/linear model in (5) can be seen as an approximation of the Box-Cox by interpolating its extreme end points. A natural and simple improvement of this is to narrow down the end points so that these are less extreme and fit the data better. As an example, if the true data generating process is with a Box-Cox parameter  $\lambda = 0.5$ , then using an approximation based on extreme points may not be the best choice. A better choice may be to interpolate between the following end points  $BC(x; \lambda = 0.3)$  and  $BC(x; \lambda = 0.7)$ .

The next question obviously is how to find these new extreme points. One approach could be using the damping measure  $\hat{\mu}$  from Definitions 1 and 2 which was obtained from two auxiliary linear models. As we

saw,  $\hat{\mu}$  is not exactly equal to the  $\lambda$  parameter and the difference may vary depending on the amount of damping and other aspects of the data. However, the point is that it needs not be very precise. In fact, if we narrow down the interpolation interval too much, the correlation between the end points will increase correspondingly. By choosing a semi-wide interval we introduce flexibility in the model, which will compensate for the fact that the interval may not be entirely correct.

Following this, we propose the approximation:

$$(7) \quad BC(x; \lambda) \sim aBC(x; (1 - \mu)(1 - k)) + bBC(x; \min(1, (1 - \mu)(1 + k)))$$

where  $0 < k < 1$ , and a 30% interpolation interval around  $\mu$  implies that  $k = 0.3$ . Strictly speaking, the right-hand side term in (7) could be questioned as it limits the Box-Cox parameter to 1 and is in conflict with the idea of approximating the Box-Cox by end points that are symmetric around  $\mu$ . Rather the right-hand side could be expressed as  $bBC(x; (1 - \mu)(1 + k))$  and the end point could go as high as  $1 + k$  if damping is low ( $\mu \rightarrow 0$ ). However, if the premise is to look for approximations in the unit interval, the specification in (7) turns out to be more efficient as it will maintain the linear function as base in case of low damping. If however, the modeller aims to search for functions being more aggressive than a linear term (Box-Cox parameters above unity) the right interpolation point could be increased further. In the following, we will refer to the approximation in (7) as the Box-Cox End Point (BCEP) approximation whereas the Log/Linear model in (5) will be referred to as the LL approximation.

### 3.1 Simulation experiment 1: compare the BCEP and the LL model when the data is Box-Cox

In the simulation experiments throughout the paper we will consider a positive continuous data generating process which can be explored in a regression context. Although this is a simpler model than a logit type model, functional form recovery issues are likely to carry over to the logit case as well.

First we generate a random variable  $d$  for travel distance. It is generated as a lognormally distributed variable in order to resemble the shrinking tails of a distance frequency. The distribution is based on a normal distribution  $X \sim N(\mu, \sigma^2)$  with  $d = \exp(X)$ . For simplicity we assume that cost  $x$  and distance  $d$  are identical.

We then generate different stochastic processes by letting different functions  $q(d)$  operate on  $d$ . If we seek a Box-Cox data generating process,  $(d) = a \frac{d^\lambda - 1}{\lambda}$ . The distribution is chosen to be largely similar to what can be observed in the Danish NTM for commuter trips. We have included trip length in the range up to 400 KM.

In Table 2 below we present a Monte Carlo experiment that shows the result of the BCEP and LL model on a true BC process. For each model, we draw 100 data sets randomly consisting of 10,000 observations. All of the reported statistics are averages across all 100 data sets. Also note that we report the loglikelihood even though models are not nested. However, as number of parameters and observations are identical the loglikelihood will give the same order of model fit as would the Akaike and Bayesian Information Criteria.

	% Deviation
--	-------------

$k$	$\lambda$	Model	LogL	0-100 KM	100-200 KM	200-300 KM	300-400 KM
0.3	0.1	BCEP	97032	0.01%	0.00%	-0.02%	-0.04%
		LL	50341	1.63%	0.22%	-1.70%	-4.09%
	0.3	BCEP	67486	2.99%	0.00%	-0.21%	-0.41%
		LL	35626	70.50%	0.85%	-4.17%	-9.71%
	0.5	BCEP	50368	1.39%	0.13%	-0.45%	-1.11%
		LL	27090	14.20%	1.23%	-4.67%	-11.20%
	0.7	BCEP	36873	4.72%	0.46%	-0.59%	-1.88%
		LL	21759	22.20%	1.29%	-3.60%	-8.63%
	0.9	BCEP	35991	4.57%	0.28%	-0.23%	-0.86%
		LL	22816	18.30%	0.66%	-1.36%	-3.28%
	0.5	BCEP	82260	0.07%	-0.02%	-0.09%	-0.16%
		LL	50341	1.63%	0.22%	-1.70%	-4.09%
	0.3	BCEP	55030	10.40%	0.00%	-0.72%	-1.42%
		LL	35626	70.50%	0.85%	-4.17%	-9.71%

Table 2: Assessment of performance of the BCEP approximation compared to the LL (Log/Linear) model for different values of interval  $k$  and  $\lambda$  parameter<sup>2</sup>.

The above tests show several things. Firstly, that in all cases the LL approximation is significantly worse than the BCEP model for all values of  $\lambda$ . In the simulation, the BCEP model has been formulated around the linear damping rate  $\hat{\mu}$ . It can also be seen that narrowing down the interval from 0.5 to 0.3 significantly increases the fit. Experiments not published here suggest that choosing an interval which is too narrow should be avoided.

### 3.2 Alternative Box-Cox approximation

The above methodology suggested a quite efficient way of approximating a Box-Cox process arbitrarily closely by linear-in-parameter functions. However, we would like to draw attention to an alternative approximation of the Box-Cox function which emerges as a Taylor expansion by using L'Hospital's rule in the following way;

<sup>2</sup> The loglikelihood (LogL) values for all of the continuous experiments are positive. This is because we allow the estimation of a scale parameter, which is generally found to be quite small. As a result, the loglikelihood will be composed of quite a few elements close to zero all of which will generate a positive contribution to the loglikelihood. This is completely valid and the larger the values of LogL the better Goodness-of-fit.

$$\begin{aligned}
(8) \quad BC(x; \lambda > 0) &= \beta \frac{x^\lambda - 1}{\lambda} \\
&= \beta \frac{e^{\lambda \ln x} - 1}{\lambda} \\
&= \beta \frac{1 + \sum_q \frac{\lambda^q (\ln x)^q}{q!}}{\lambda} \\
&\sim \beta \left( \frac{1}{\lambda} + \ln x + \frac{\lambda}{2} (\ln x)^2 + \frac{\lambda^2}{6} (\ln x)^3 + \dots + O(K) \right)
\end{aligned}$$

In the brackets, the Box-Cox is now represented as a series of non-negative terms (if we assume  $x > 1$ ) all scaled with positive parameters. When multiplied with a negative scale parameter  $\beta$ , it becomes a simple polynomial expression with only negative terms  $\beta_0, \beta_1, \beta_2, \dots$  which can easily be estimated (refer to equation (7) below). A challenge here is that the log-power terms are correlated. Although excessive correlation should always be avoided, it may not in itself be a problem to have correlated terms if we are operating on a single variable such as a generalised travel time (the log and linear terms are correlated as well). Even so, to reduce the correlation issue we suggest a specification that results from the following steps:

- i) Initially, include a sufficient number of terms to cope with the non-linearity.
- ii) Remove positive terms one by one until only negative terms exist.
- iii) Use backward selection (remove the least significant terms first) until only two terms are included.

These steps represent a search process for a parsimonious two-parameter linear model. As we will see in a subsequent simulation test these models perform very well in explaining not only Box-Cox but also other types of data generating processes. We will refer to this model as the LP model (Log-Power model) as restated in (9) below

$$(9) \quad LP(x) = \beta_0 + \beta_1 \ln x + \beta_2 \ln x^2 + \beta_3 \ln x^3 \dots$$

If there is a minimum of cost damping the resulting model in (9) will generally include log-terms raised to a high power, e.g. 7-9. This makes the down-testing of the model somewhat cumbersome (may need 6-9 model runs). What we therefore suggest, if damping is low (which can be tested by examining the linear damping rate in Definition 2) we therefore suggest including a linear term as well and then test the model down to include only one additional log-power term. We will refer to this model as the LLP model (Linear Log-Power model) as shown in (10) below.

$$(10) \quad LLP(x) = \beta_0 + \beta_1 x + \beta_2 \ln x + \beta_2 \ln x^2 + \beta_2 \ln x^3 \dots$$

The down-testing of the LLP model should be carried out in a similar manner as for the LP model except that the linear term should always be included. In other words, it is a matter of testing whether the second term should be  $\ln x^2, \ln x^3$  or  $\ln x$  raised to an even higher power.

Yet a different model is a model which includes a product of the linear and logarithmic term alongside the linear term. This model is shown in (11) below and we will refer to this model as the xL model

$$(11) \quad xL(x) = \beta_0 + \beta_1 x + \beta_2 x \ln x$$

The slope of the xL model is typically rather different from the one of the LP and LLP models. It is common to have a large negative  $\beta_1$  and a smaller positive  $\beta_2$ . This generally means that the micro-economic condition that cost has a negative derivative is not fulfilled as  $x \rightarrow \infty$ . From an implementation perspective it is necessary to consider the “domain of validity” of the model. Hence, we need to examine where  $\frac{\partial xL(x)}{\partial x} < 0$ , which in turn (if  $\beta_2 > 0$  and  $\beta_1 < 0$ ) implies that  $\beta_2(\ln x + 1) < \beta_1 \Rightarrow x < \exp\left(\frac{\beta_2}{\beta_1} - 1\right)$ . In all of the tests (simulation as well as real-world examples) where damping has been moderate to low this condition has been valid for the entire observed sample and quite a bit beyond. Below we explore the various alternative models in a series of Monte Carlo experiments.

### 3.3 Simulation experiment 2: Comparison of BCEP, LL, LLP, LP and xL when the data is Box-Cox

The experiment reported below in Table 3 is based on the true data being Box-Cox.

$\lambda$	Model	LogL	% Deviation			
			0-100 KM	100-200 KM	200-300 KM	300-400 KM
0.2	BCEP	79359	0.09%	-0.01%	-0.09%	-0.17%
0.2	LL	41478	3.80%	0.49%	-3.09%	-7.44%
0.2	LLP	NA	NA	NA	NA	NA
0.2	LP ( $\ln x + \ln x^2$ )	54146	-1.02%	0.37%	1.25%	2.00%
0.2	xL	29412	0.38%	-5.03%	5.48%	22.66%
0.5	BCEP	50368	1.39%	0.13%	-0.45%	-1.11%
0.5	LL	27090	14.15%	1.23%	-4.67%	-11.23%
0.5	LLP ( $x + \ln x^2$ )	42805	-3.15%	0.60%	-0.48%	-2.26%
0.5	LP ( $\ln x + \ln x^4$ )	40245	5.47%	-0.37%	-0.88%	-0.97%
0.5	xL	28542	-0.10%	-2.51%	2.01%	8.85%
0.8	BCEP	34664	7.49%	0.46%	-0.46%	-1.48%
0.8	LL	20816	33.15%	1.16%	-2.64%	-6.04%
0.8	LLP ( $x + \ln x^3$ )	34554	-13.59%	0.09%	0.18%	-0.11%
0.8	LP ( $\ln x^2 + \ln x^5$ )	29078	-10.55%	-0.79%	1.17%	3.24%
0.8	xL	35578	-0.47%	-0.49%	0.20%	1.20%

Table 3: Assessment of performance of the BCEP, LL, LP and LLP models for approximating Box-Cox for different values of  $\lambda$  and with  $k = 0.3$ .

If judged by the loglikelihood function there is a tendency that the BCEP model is the better model in all experiments except for  $\lambda = 0.8$  where the xL performs exceptionally well. However, if we look at the performance in the tail of the distribution we will see that the LP and LLP models perform better for some of the values of  $\lambda$ . In fact, from a practical modelling point of view this may be more important as capturing the first part of the curve is not the essential challenge (other variables commonly available in travel data may help explaining this part of the curve). We also see that the performance of the LL model is not very good.

The above experiment was based on the premise that the data generating process was Box-Cox. Below we will consider a case where the data is not Box-Cox to get an idea of the robustness of the different forms.

### 3.4 Simulation experiment 3: Comparison of Box-cox, BCEP, LL, LLP, LP and xL when the data is not Box-Cox

We will explore how well the different models can recover functional forms that are not Box-Cox but still monotone and marginally decreasing. In particular, we will compare the Box-Cox approximations with a true Box-Cox model estimated by full information maximum likelihood.

We consider three data generating processes

$$(12) \quad y_{\log/\ln}(x) = a_0 + a_1x + a_2\ln x$$

$$(13) \quad y_{x/x\ln x}(x) = a_0 + a_1x + a_2x\ln x$$

$$(14) \quad y_{\ln 2x/x\ln x}(x) = a_0 + a_1(\ln x)^2 + a_2x\ln x$$

Results are given in Appendix A, however, below we outline the main findings.

- For the LL data the true Box-Cox model is not a good model. In that sense, the “coin is two-sided” in that the LL model is weak when confronted with a Box-Cox data and that the Box-Cox model is weak when confronted with a LL data.
- Only in 1 out of the 9 experiments the Box-Cox model turns out to be the better model.
- In 6 out of 9 experiments the BCEP model outperforms the Box-Cox model. This is because it uses its more flexible end points to adjust the function to the data. It is not always the case though as it depends on the data.
- The amount of damping is a very important determinant for the success of the various functions. Generally, if damping is high, the LP model is a natural choice. If on the other hand damping is low, the LLP or xL models seem to be good candidates as well.

A conclusion based on the specific simulation experiments is that the researcher in search for functional forms to explain the disutility of increasing travel time and cost should be careful not to focus on a particular type of function. Many function classes are needed in the tool-box. The results suggest that Box-Cox is generally not necessarily very robust to deviations from the assumption of Box-Cox.

We also note (as previously stated in connection with Definition 2) that the linear damping rate is slightly higher than the corresponding Box-Cox damping rate irrespectively of the data generating process. The linear damping estimate is useful for choosing among the different function types, among which certain types favour data which is highly damped and others data which is damped, only to a little extent. The linear damping estimate (if corrected for its upward bias) may also be a fairly good guess of the Box-Cox parameter without the need for running a non-linear model or it may be used as initial parameter guess prior to estimation.

## 4 Real-world tests based on NTM data

To test the performance of the various functional forms and the relative impact on elasticities we have applied the functional forms to data used for the NTM. For simplicity we will only consider weekday commuter trips. The data consist of 22,761 revealed preference observations from the national Danish Travel Survey.

The structure of the demand model is consistent with what has been applied in the NTM. That is, a nested logit structure with choice of destination at the upper level and choice of mode at the lower level.<sup>3</sup> For simplicity, we have left out the remaining part of the model involving frequency choice and car ownership. For the choice of destination we have used sampling of alternatives (60 out of 907 are sampled) based on a stratified importance sampling setup. The models are corrected for sampling errors as suggested in Guevara and Ben-Akiva (2013). This will provide consistent estimates with the applied nesting structure. Transport modes include walk, bike, car, car as passenger, and public transport.

The destination-mode choice model is fairly comprehensive and includes around 80 parameters. Variables included:

- Generalised time variables
- Regional dummies
- Additional public transport level-of-service variables
- Station proximity (at both ends), connector time, walk time,...
- Socio-economic variables such as gender, age and income.
- Car availability
- Size variable related to the choice of destination based on employment in 127 sectors

The paper does not allow a detailed description of the model and all of the underlying variables.<sup>4</sup>

### 4.1 Model structure

In the following, let  $d$  represent destination (907 in total) and  $i$  income intervals (currently 10 income bands for gross personal income). Function  $1_i$  is an indicator function for individual  $n$  belonging to income group  $i$ . Monetary costs for all modes are deflated by an income dependent exogenous VoT estimate  $VoT(i)$ . For simplicity we skip the index representing individuals.

In the estimation process the following cost functions are applied for all models.

$$(15) \quad cost_{car}(d) = c_{car}(d) \left[ \sum_i \frac{1_i}{VoT(i)} \right]$$

$$(16) \quad cost_{carp}(d) = 0$$

---

<sup>3</sup> The nesting structure varies in the NTM in the sub-models for the different trip purposes. However, for commuting this was the best nesting structure statistically speaking.

<sup>4</sup> Documentation for the NTM is available from the first author upon request.



$$(17) \quad cost_{pub}(d) = c_{pub}(d) \left[ \sum_i \frac{1_i}{VoT(i)} \right]$$

where  $c_{car}(d)$  is the car cost to destination  $d$ .

Time for cars (and passengers) as shown in (18) is a combination of free-flow time  $fft_{car}$ , congestion time  $ct_{car}$ , ferry-sailing-time  $fst(d)$  and ferry-waiting-time  $fwtd$ . All of these attributes are calculated in an assignment model.

$$(18) \quad time_{car}(d) = [fft_{car}(d) + \gamma_1 ct_{car}(d)] + fst(d) + \gamma_2 fwtd$$

Public transport time (calculated in a schedule-based assignment model) consists of three weighted components, vehicle-time  $Inv_{pub}(d)$ , number of shifts  $ns_{pub}$  and waiting time  $wt_{pub}$ .

$$(19) \quad time_{pub}(d) = Inv_{pub}(d) + \vartheta_1 ns_{pub} + \vartheta_2 wt_{pub}$$

In the model, time and cost are joined in a generalised time measure  $GTT_m$

$$(20) \quad GTT_m(d) = time_m(d) + cost_m(d)$$

To account for differences between the transport market in and around Copenhagen and the rest of the country we allow for a separate parameterisation for Copenhagen. To reduce the complexity of the analysis we only use functional forms composed of two functions  $f$  and  $g$ , each scaled by a corresponding parameter  $\beta_{1,r}$  and  $\beta_{2,r}$  (similar to what was used in the simulation experiments). Moreover, we apply the same functional form for Copenhagen ( $r = 1$ ) and the rest of Denmark ( $r = 2$ ) although the parameters will be different.  $1_r$  is the regional dummy function, which is one if region  $r$  is chosen and zero elsewhere. This setup leads to the following utility functions for the various linear approximations of GTT

$$(21) \quad V_m(d) = \sum_{r \in \{1,2\}} 1_r \left( \beta_{1,r} f(GTT_m(d)) + \beta_{2,r} g(GTT_m(d)) \right)$$

In case of the Box-Cox model, which is non-linear, the form becomes

$$(22) \quad V_m(d) = \sum_{r \in \{1,2\}} \beta_r 1_r BC(GTT_m(d), \lambda)$$

## 4.2 Damping and overall goodness-of-fit

In the functional form analysis for the NTM (refer to setup above in Section 4.1) we start by assessing the damping degree for the two segments in the model.

NTM test	Copenhagen (r=1)	Outside Copenhagen (r=2)
$\hat{\lambda}$	0.691	0.753
$1 - \hat{\mu}$	0.733	0.896

Table 4: Estimated Box-Cox parameter  $\hat{\lambda}$  and  $1 - \hat{\mu}$  in the NTM.

The exact Box-Cox estimate was found by a combination of a bi-section algorithm and Monte Carlo simulation. Results indicate which was seen in the simulation, that  $\hat{\lambda} < 1 - \hat{\mu}$ . For  $r = 1$  (Copenhagen) the relative difference largely follows what was found in the simulation. For  $r = 2$  (outside Copenhagen) the difference is higher. This indicates not only a different damping in the two segments but also a somewhat different functional form.

The overall goodness-of-fit is presented below in Table 5.

Model	Description	LogL	Consistent with RUM
BC		-73661	☑
BCEP	$k = 0.3$	-73679	☑
LL	$x + \ln x$	-73790	☑
LLP	$x + \ln^2 x$	-73738	☑
	$x + \ln^3 x$	-73667	☑
LP	$\ln x + \ln^5 x$	-73665	No
	$\ln^2 x + \ln^5 x$	-73640	No
xL	$x + x \ln x$	-73639	On the sample ☑, not $\forall x > 0$

Table 5: Goodness-of-fit of the different models based on data from the NTM.

The first thing to note is that different functional forms cause huge shifts in the loglikelihood score. Exactly 151 log-likelihood points separate the best (xL) and worst (LL) model, which is more than the increase of going from a multinomial logit to the nested logit. Another thing to note is that the non-linear Box-Cox model actually performs very well. We also see that it is very well approximated by the BCEP model. If the interpolation interval represented by  $k = 0.3$  was narrowed down further (refer to equation (7) in Section 3) the BC model and the BCEP model would have had identical performance. In fact, as we shall see later the comparison between the two models in terms of elasticities and ability to replicate the distance distribution of elasticities is so similar that we will only consider the BCEP model in the remainder of the paper (Section 4.4). Another comment is that, as the damping percentage is relative low, the LP model is not really suitable for these data. As a result, it produces models which are not entirely consistent with RUM for the whole domain of GTT. This causes a positive parameter for one of the log-power terms, which in turn causes a violation of the first property in Section 1.

### 4.3 Elasticities and distance distribution recovery

Besides looking at the overall fit of the model it is important and relevant to assess whether the difference in goodness-of-fit corresponds to differences in the sensitivity of the model. Generally, if we look at average elasticities over the sample the differences are relative small. As a result, we are doing well “on average” irrespective of which method we consider and we will not consider those further. The differences are in the tails of the distribution and this is important to consider as a considerable part of the total mileage is captured in the tails. The focus on the tails is relevant in many forecast situations where large infrastructure projects are analysed as these usually have a high proportion of long-distance trips. Danish examples are The Great Belt Bridge (1998), the Øresund Bridge (2000), and the Fehmarn Belt Fixed Link (to open in 2020). Below in Table 6 we consider average elasticities for trips over 50 kilometres and the percentage divergence of these elasticities when compared with the best model (the xL model) in Table 7.

Model	Description	Elast.			
		Car cost	Car time	Pub cost	Pub time

BC		-1.248	-1.023	-1.242	-1.729
BCEP	$k = 0.3$	-1.250	-1.023	-1.247	-1.735
LL	$x + \ln x$	-1.362	-1.149	-1.323	-1.802
LLP	$x + \ln^3 x$	-1.210	-0.980	-1.217	-1.695
xL	$x + x \ln x$	-1.185	-0.955	-1.186	-1.664

Table 6: Direct elasticities for car and public transport in the tail of the distribution (>50 KM).

Model	Car cost	Car time	Pub cost	Pub time
BC	5.32%	7.12%	4.72%	3.91%
BCEP	5.49%	7.12%	5.14%	4.27%
LL	14.94%	20.31%	11.55%	8.29%
LLP	2.11%	2.62%	2.61%	1.86%

Table 7: Overprediction of elasticities compared to the best model (xL) in the tail of the distribution (>50 KM).

As seen, the LL model overpredicts the sensitivity in the tails of the distribution compared to the best model. Clearly, this is not the same as to say that the LL model overpredicts compared to the true model so Table 7 offers nothing more than a comparison relative to the xL model. The overprediction varies between 8% and 20%. Moreover, the BC and the BCEP models also seem to slightly overpredict sensitivity in the range of 4%-7%.

When considering the distance recovery of the model, which is basically a misspecification test, it reveals certain problems in the tails for both public transport and car as seen in Table 8.

Model	Description	%deviation car-tail	%deviation pub-tail
BC		1.57%	-11.55%
BCEP	$k = 0.3$	2.29%	-11.55%
LL	$x + \ln x$	-1.38%	-16.87%
LLP	$x + \ln^3 x$	4.66%	-15.03%
LP	$\ln^2 x + \ln^5 x$	-7.96%	-18.51%
xL	$x + x \ln x$	3.47%	-15.13%

Table 8: Deviation between shares of observed and modelled trips above 50 KM.

This means that the comparison in section 4.2 was based on a slightly misspecified model. We have kept this in the paper as it provides an examination of how well the different functional forms perform under a misspecified and a more correctly specified model to be introduced below in section 4.4.

#### 4.4 Results from an improved model

Due to the previous misspecification an improved model was considered. For this model we only consider four functional forms: The BCEP, LL, LLP and xL models. The reader will notice that due to the low degree of damping all of these functions include a linear term in their specifications. The BCEP model includes the linear term because we apply a 30% interval, which in this case hits the upper limit of 1. The way the specification was improved compared to section 4.1 was to simply split the estimation of the public transport GTT (other modes are not changed) into cost and time as seen in (23) below

$$(23) \quad V_{pub}(d) = \sum_{r \in \{1,2\}} \beta_{1,r} cost_{pub}(d) + \beta_{1P,r} time_{pub}(d) + \beta_{2,r} f(GTT_m(d))$$

So by this re-construction we keep the cost-parameter constant across alternatives but let the travel-time vary. Due to this we implicitly estimate a VoT (or a scaling of the VoT) for public transport. Below in Table 9 we show the overall model fit.

Model	Description	LogL	Consistent with RUM
BCEP	$k = 0.3$	-73665	☑
LL	$x + \ln x$	-73765	☑
LLP	$x + \ln^2 x$	-73667	☑
	$x + \ln^3 x$	-73639	☑
xL	$x + x \ln x$	-73612	On the sample ☑, not $\forall x > 0$

Table 9: Model fit of the various improved models based on data from the NTM.

By comparing these results with results based on the misspecified model we can see that all models improve their loglikelihood value. The highest absolute increase is in the LLP (the  $x + \ln^2 x$  model) and the second highest is in the xL model. The smallest increase is in the BCEP model. Below in Table 10 we look at the recovery of the distance distribution in the new improved model. For the LLP models, we present only the results for the best model.

Model	%deviation car-tail	%deviation pub-tail
BCEP	0.3%	-3.07%
LL	-4.78%	-3.07%
LLP	1.43%	-1.06%
xL	0.08%	-1.43%

Table 10: Deviations between shares of observed and modelled trips above 50 KM. Improved model.

As we can see, there is a very significant improvement in the fit for the tails. For the xL model the fit for cars is almost perfect, whereas for public transport there is less than 2% deviation which for most standards is acceptable.

If we look at the elasticities in Table 11 these are shifted downwards in the improved model.

Model	Elast.			
	Car cost	Car time	Pub cost	Pub time
BCEP	-1.263	-1.035	-1.242	-1.609
LL	-1.381	-1.169	-1.323	-1.635
LLP	-1.230	-0.999	-1.219	-1.517
xL	-1.206	-0.974	-1.181	-1.466

Table 11: Direct elasticities for car and public transport in the tail of the distribution (>50 KM) for the improved model.

If we look at the relative overprediction of the different models compared to the xL model these have generally not improved and seem more or less unchanged.

Model	Car cost	Car time	Pub cost	Pub time
BCEP	4.73%	6.26%	5.17%	9.75%

LL	14.51%	20.02%	12.02%	11.53%
LLP	1.99%	2.57%	3.22%	3.48%

Table 12: Overprediction of elasticities compared to the best of the improved models (xL) in the tail of the distribution (>50 KM).

The results show that the choice of functional form has great impact on the tail elasticities. This may be important if the underlying model is used for assessment of infrastructure where a majority of the trips are long-distance trips.

## 5 Conclusion and summary

This paper has investigated functional form in transport demand modelling and in particular the issue of cost damping and how choosing a proper functional form will significantly influence the goodness-of-fit, elasticities, and the recovery of the revealed distance distribution. More specifically the paper contributes to the literature with the following findings.

Firstly, we point out a way of determining the amount of cost damping based on two auxiliary linear models. The assessment of the amount of cost damping is shown to be essential to the choice of functional form. If damping is low, models with a linear cost as base are generally recommended. On the other hand, if damping is high, models with a logarithmic cost as base should generally be used.

Secondly, we consider approximations of the Box-Cox model and show that good approximations can be obtained by either a Box-Cox End Point (BCEP) approximation or by a Taylor expansion of a slightly rewritten Box-Cox form. In the BCEP model we approximate the Box-Cox by first estimating the damping rate of the underlying process and then subsequently fitting the Box-Cox curve by interpolating between Box-Cox end points defined in some proper region of the approximated Box-Cox parameter. For the Taylor expansion we show that the Box-Cox works out as a sequence of log-power terms. Subsequently, based on the Taylor expansion, we argue that by adding a linear term and a product of a linear and a logarithmic term two potentially interesting models arise which are relevant for low-damping processes.

Thirdly, we analyse the various functional forms in several data generating processes and conclude based on simulation experiments that the Box-Cox model is not particularly robust. In fact, the BCEP counterpart generally has better model fit when the data is not a Box-Cox. This is because it uses its flexible end points as a mean of correction.

When tested on real-world data the various functional forms turn out to produce huge shifts in the log-likelihood score as well as significant differences in the elasticities in the tail of the distance distribution. In the real-world example damping was shown to be relatively low. In our application, it turned out that the xL model had significantly better model fit than all the other models including the non-linear Box-Cox model. The worst model in terms of fit was the Log/Lin model. Results were robust whether they were based on a slightly misspecified model or an improved model. The final implementation of the Danish National Model benefitted from the findings in the paper, and the log-power function was widely used in the demand framework across segments.

The research on functional form seems to be a widely uncultivated field except for contributions summarised by Gaudry (2010) and Daly (2010). This seems unjustified in light of its large impact on model

fit and model output as also documented in this paper. Certainly more research is needed in this field. Particularly, it would be relevant to investigate how VoT is recovered across relevant functional forms as many of the recommended forms may imply correlation issues. The recovery of VoT is obviously a tougher task than recovering a generalised travel time form as considered in this paper.

## Acknowledgment

The research in the current paper was partly funded by the Danish Ministry of Transport during the work with the new Danish National Transport Model and partly by the ACTUM project financed by the Danish strategic research council.

We would like to thank Andrew Daly for fruitful discussions and for sharing insightful best practise experience with us. We would also like to thank Marc Gaudry for a number of very useful comments and references. We also like to thank several people in the audience at the hEART Conference (Stockholm, 2013 September 4-6) and at the UNITE Conference (Copenhagen, 2013 September 14-18) for good suggestions and recommendations. Finally, we would also like to thank the three anonymous reviewers for well-placed comments and suggestions.

## References

- Ben-Akiva M, Daly AJ, Gunn HF (1987) Destination Choice Models: Design and Appraisal. Proceedings of the 10th PTRC Summer Annual Meeting, University of Warwick
- Ben-Akiva M, Lerman S (1985) Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge
- Beser MH, Algers S (2002) SAMPERS – The New Swedish National Travel Demand Forecasting Tool. In Lundqvist, L., Mattsson, L.-G., (eds), National Transport Models – Recent Development and Prospects, Advances in Spatial Science, Springer-Verlag, Berlin, Ch. 9:101-118.
- Börjesson M (2010) Forecasting demand for high-speed rail. Proceedings of the 38<sup>th</sup> European Transport Conference, Glasgow, Scotland
- Daly A (2012) Two approaches to non-linear transformations, Unpublished slides presented 12<sup>th</sup> October 2012, Cambridge, Rand Europe.
- Daly A (2010) Cost damping in travel demand models. Technical report, Rand Europe, see [http://www.rand.org/pubs/technical\\_reports/TR717.html](http://www.rand.org/pubs/technical_reports/TR717.html) (accessed 15-12-12)
- Daly A, Carrasco J (2009) The influence of trip length on marginal time and money values. In The Expanding Sphere of Travel Behaviour Research: Selected papers from the proceedings of the 11<sup>th</sup> Conference on Travel Behaviour Research, Kitamura, R. et al. (eds.), Emerald Books

- Gaudry M (2010) Quebec-Windsor Corridor High Speed Rail Market Forecast Profiles in Context: Level-of-Service Response Curvature Sensitivity and Attitude to Risk or to Distance in Forty Logit Core Model Applications of the Law of Demand, Agora Jules Dupuit, AJD-127, [www.e-ajd.org](http://www.e-ajd.org)
- Gaudry M, Wills M (1978) Estimating the Functional Form of Travel Demand Models. *Trans. Research* 12, pp. 257-289
- Gaudry M, Jara-Diaz S, Ortuzar JD (1989) Value of time sensitivity to model specification. *Transportation Research B* 23(2):151-158
- Gaudry M, Blum U, Liem T (2000) Turning Box-Cox, including Quadratic Forms in Regression. In Gaudry, M. and S. Lassarre, (eds), *Structural Road Accident Models: The International DRAG Family*, Pergamon, Elsevier Science, Oxford, Ch. 14:335-346
- Guevara CA, Ben-Akiva M (2013) Sampling of alternatives in Multivariate Extreme Value (MEV) models, *Transportation Research B* 48:31-52
- Hensher D, Johnson L (1981) *Applied Discrete Choice Modelling*. Croom Helm, London
- Hyman G (2007) Three propositions relating to the realism of car traffic elasticities. Department for Transport, unpublished
- Lapparent M, de Palma A (2002) Nonlinearities in the Valuations of Time Estimates. Proceedings of the 30<sup>th</sup> PTRC Annual Meeting, PTRC, London
- Mandel B, Gaudry M, Rothengatter W (1994) Linear or Nonlinear Utility Function in Logit Models? The Impact on German High-Speed Rail Demand Forecast. *Transportation Research B*, 28(2):91-101
- Jara-Diaz, S (2007) *Transport Economic Theory*. Emerald Group Publishing Limit
- Pinjari A, Bhat C, (2006) Nonlinearity of response to level of service variables in travel mode choice models. *Transportation Research Record, Journal of the Transportation Research Board, National Academy of Sciences* 1977, pp. 67-74
- Rich J, Mabit S (2012) A long-distance travel demand model for Europe. *European Journal of Transport and Infrastructure Research* 12(1):1-20
- Rohr C, Fox J, Daly A, Patruni B, Patil S, Tsang F (2010) Modelling long-distance travel in the UK. The 38<sup>th</sup> European Transport Conference, Glasgow, Scotland
- Wardman M (2008) Is Generalised Cost Justified? Proceedings of the 36<sup>th</sup> ETC conference, Noordwijkerhout, Netherlands
- WSP (2011) Höghastighetståg – modellutveckling. Forskningsrapport, WSP, [http://fudinfo.trafikverket.se/fudinfoexternwebb/Publikationer/Publikationer\\_001401\\_001500/Publikation\\_001404/HHT%20rapport\\_110622.pdf](http://fudinfo.trafikverket.se/fudinfoexternwebb/Publikationer/Publikationer_001401_001500/Publikation_001404/HHT%20rapport_110622.pdf)

## Appendix A: Acronyms

Acronym	Equation	Description
Box-Cox	(4)	True Box-Cox form as shown in equation (4). Is also referred to as BC.
LL	(5)	Log-Linear form ( $x + \ln x$ )
BCEP	(6)	Box-Cox End Point form ( $aBC(x; (1 - \mu)(1 - k)) + bBC(x; \min(1, (1 - \mu)(1 + k)))$ )
LLP	(10)	Linear-Log-Power form ( $x + \ln^q x; q \in \{2, 3, \dots, Q\}$ )
LP	(9)	Log-Power form ( $\ln^{q_1} x + \ln^{q_2} x; q_1, q_2 \in \{1, 2, \dots, Q\} q_1 < q_2$ )
xL	(11)	Linear plus linear and log multiplicative form ( $x + x \ln x$ )

Table 13: Short names for the various functional forms.

## Appendix B: Simulation experiments when the data is not Box-Cox.

The parameters  $\hat{\lambda}$  and  $\hat{\mu}$  represent the Box-Cox and linear damping rate of the underlying data generating process.

### Recovery of the LL model

First we consider recovery of the LL model given by

$$(24) \quad y_{\ln/\ln}(x) = a_0 + a_1 x + a_2 \ln x.$$

These results are presented in Table 14.



Model	$\hat{\lambda}$	$\hat{\mu}$	LogL	% Deviation			
				0-100 KM	100-200 KM	200-300 KM	300-400 KM
BC	0.293	0.369	18746	-0.82%	0.02%	4.87%	10.12%
BCEP			19094	-0.86%	-0.19%	4.31%	9.29%
LLP			14403	-2.00%	-1.02%	3.56%	8.65%
LP			24263	-0.35%	-0.39%	2.92%	6.97%
BC	0.556	0.613	16085	-1.48%	-1.13%	4.06%	8.88%
BCEP			17376	-1.37%	-1.10%	3.23%	7.29%
LLP			17320	-1.59%	-0.80%	2.28%	4.87%
LP			25378	-0.22%	-0.99%	1.89%	5.22%
BC	0.804	0.826	17113	7.95%	-1.25%	1.73%	4.49%
BCEP			19930	6.71%	-0.82%	1.29%	3.10%
LLP			21370	8.00%	-0.42%	1.03%	2.09%
LP			18457	0.69%	-1.53%	2.55%	6.71%

Table 14: Evaluation of the various functional forms for a LL data generating process. “Tails” represent the 5% longest trips.

### Recovery of the xL model

Below we consider recovery of the xL model.

$$(25) \quad y_{x/x \ln x}(x) = a_0 + a_1 x + a_2 x \ln x$$

Results are presented in Table 15. This is a quite different type of model and does not collide with any of the suggested models except of course for the xL model itself.

Model	$\hat{\lambda}$	$\hat{\mu}$	LogL	% Deviation			
				0-100 KM	100-200 KM	200-300 KM	300-400 KM
BC	0.661	0.733	26860	1.84%	1.21%	-1.59%	-5.69%
BCEP			23512	2.89%	1.68%	-1.99%	-7.25%
LL			15589	4.89%	2.59%	-5.84%	-16.00%
LLP			21620	2.68%	2.24%	-2.47%	-9.26%
LP			35295	0.66%	0.32%	-0.78%	-2.74%
BC	0.796	0.846	34369	0.98%	0.49%	-0.40%	-1.63%
BCEP			27156	2.12%	0.94%	-0.84%	-3.16%
LL			17817	3.98%	1.59%	-3.21%	-8.24%
LLP			23849	2.15%	1.36%	-1.35%	-4.77%
LP			35687	-0.75%	-0.34%	0.56%	1.48%
BC	0.898	0.925	48416	0.34%	0.13%	-0.07%	-0.34%
BCEP			35600	1.23%	0.42%	-0.32%	-1.22%
LL			24740	2.70%	0.81%	-1.51%	-3.73%
LLP			30771	1.43%	0.69%	-0.63%	-2.16%
LP			23174	-2.93%	-1.28%	1.79%	5.00%

Table 15: Evaluation of the various functional forms for a xL data generating process. “Tails” represent the 5% longest trips.

As seen from  $\hat{\lambda}$  and  $\hat{\mu}$  this type of data is not naturally a high-damping process. For the high to medium high damping data generating processes, the LP model performs well. For low damping the BC model is the better model.

## Recovery of $\log^2 x/x \log x$ model

Below we consider recovery of data of the form;

$$(26) \quad y_{\ln 2x/x \ln x}(x) = a_0 + a_1(\ln x)^2 + a_2 x \ln x.$$

Results are presented in Table 16.

Model	$\hat{\lambda}$	$\hat{\mu}$	LogL	% Deviation			
				0-100 KM	100-200 KM	200-300 KM	300-400 KM
BC	0.198	0.27	20801	0.38%	0.15%	5.26%	10.66%
BCEP			20868	0.38%	-0.06%	4.85%	10.13%
LL			38584	0.03%	-0.23%	0.63%	1.78%
LLP			13822	0.82%	-1.47%	4.66%	11.53%
LP			19154	0.51%	0.95%	7.06%	13.12%
BC	0.362	0.437	15519	0.55%	-0.57%	6.22%	13.12%
BCEP			15928	0.50%	-0.85%	5.40%	11.90%
LL			31662	0.07%	-0.41%	1.01%	2.72%
LLP			13480	0.80%	-1.48%	3.96%	9.61%
LP			19045	0.15%	-0.78%	4.77%	10.85%
BC	0.875	0.862	8836	-0.08%	-2.19%	2.27%	6.71%
BCEP			10138	0.16%	-1.82%	2.22%	6.06%
LL			15589	0.39%	-1.04%	1.95%	4.73%
LLP			8355	-0.27%	-2.10%	1.79%	5.55%
LP			5883	0.36%	-2.76%	5.00%	12.40%

Table 16: Evaluation of the various functional forms for a  $\log^2 x/x \log x$  data generating process. “Tails” represent the 5% longest trips.